

Kernel Target Alignment Parameter: A New Modelability Measure for Regression Tasks

Marcou G., Horvath D., Varnek A.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2015 American Chemical Society. In this paper, we demonstrate that the kernel target alignment (KTA) parameter can efficiently be used to estimate the relevance of molecular descriptors for QSAR modeling on a given data set, i.e., as a modelability measure. The efficiency of KTA to assess modelability was demonstrated in two series of QSAR modeling studies, either varying different descriptor spaces for one same data set, or comparing various data sets within one same descriptor space. Considered data sets included 25 series of various GPCR binders with ChEMBL-reported pKi values, and a toxicity data set. Employed descriptor spaces covered more than 100 different ISIDA fragment descriptor types, and ChemAxon BCUT terms. Model performances (RMSE) were seen to anticorrelate consistently with the KTA parameter. Two other modelability measures were employed for benchmarking purposes: the Jaccard distance average over the data set (Div), and a measure related to the normalized mean absolute error (MAE) obtained in 1-nearest neighbors calculations on the training set ($\text{Sim} = 1 - \text{MAE}$). It has been demonstrated that both Div and Sim perform similarly to KTA. However, a consensus index combining KTA, Div and Sim provides a more robust correlation with RMSE than any of the individual modelability measures.

<http://dx.doi.org/10.1021/acs.jcim.5b00539>
